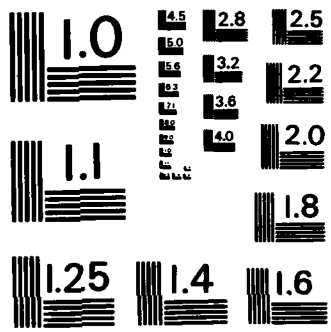
AD-A121 471 ON THE STABILITY OF BAYES ESTIMATORS FOR GAUSSIAN PROCESSES(U) FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS I W MCKEAGUE SEP 82 ARO-19367.1-MA 1/1

UNCLASSIFIED DAAG29-82-K-0168 F/G 17/2 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(12)

AD A121471

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 19367.1-MA	2. GOVT ACCESSION NO.: AD-A121471
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): On the Stability of Bayes Estimators for Gaussian Processes	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Ian W. McKeague	8. CONTRACT OR GRANT NUMBER(s): DAAG29 82 K 0168
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Florida State University, Tallahassee, FL 32306	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE: Sep 82
	13. NUMBER OF PAGES: 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE NOV 9 1982 S D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Gaussian processes
Gaussian noise
estimators
Bayes theorem
signal processing
electromagnetic noise
DELTA 0

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider the Bayes estimator δ_0 for a Gaussian signal process observed in the presence of additive Gaussian noise under contamination of the signal or noise by QN-laws. ~~introduced by Gualtierotti (1979).~~ Upper bounds on the increase in the mean square error of δ_0 over the minimum possible mean square error under contaminated noise or contaminated signal are given. It is shown that the performance of δ_0 is relatively close to optimal for small amounts of contamination.

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

On the Stability of Bayes Estimators for Gaussian Processes

by

Ian W. McKeague

FSU Statistics Report M642
USARO Report No. D-56

September, 1982

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Research supported by a Florida State University COFRS Grant and US Army Research Office Grant DAAG29-82-K-0168.

Key Words: Robustness, Bayes estimators, Gaussian processes, QN-laws.

AMS (1980) Subject Classification: Primary 62F35, 62M20; Secondary 62F15, 60G35, 60B11.

On the Stability of Bayes Estimators for Gaussian Processes

by

Ian W. McKeague

Abstract

We consider the Bayes estimator δ_0 for a Gaussian signal process observed in the presence of additive Gaussian noise under contamination of the signal or noise by QN-laws, introduced by Gualtierotti (1979). Upper bounds on the increase in the mean square error of δ_0 over the minimum possible mean square error under contaminated noise or contaminated signal are given. It is shown that the performance of δ_0 is relatively close to optimal for small amounts of contamination.

1. Introduction.

The Bayesian approach to the robust estimation of a signal in the presence of noise has been studied extensively in recent years. Some authors, including Blum and Rosenblatt (1967), Solomon (1972), Watson (1974) and Berger (1982) have discussed procedures which can be used when only vague information concerning the prior distribution is available. Others, including Box and Tiao (1968), Masreliez (1975) and Ershov and Liptser (1978) have constructed estimators which are robust with respect to contamination of the noise distribution.

The purpose of the present article is to study the performance of the usual Bayes estimator (denoted δ_0) for Gaussian prior and additive Gaussian noise under certain deviations from normality in either the prior or the noise distribution. It is shown that the performance of δ_0 is relatively close to optimal for small amounts of contamination. The main results of the paper give upper bounds on the increase in the mean square error of δ_0 over the minimum possible mean square error under a specific contaminated prior or contaminated noise distribution. These results make it possible to assess the loss caused by the use of δ_0 under non-Gaussian conditions. The contaminated Gaussian laws used in this paper are QN-laws (quasi-noise laws) which were introduced by Gualtierotti (1979). QN-laws form an appropriate class of contaminated Gaussian laws for some infinite dimensional models arising in communication theory (see Gualtierotti, 1980). Gualtierotti (1982) recently studied the stability of signal detection under mixtures of Gaussian laws as well as QN-laws. Contamination by Gaussian mixtures was shown to lead to worse behavior than contamination by QN-laws. In the present paper attention is restricted to contamination by QN-laws.

Section 2 contains some preliminary material on measures on locally convex spaces and a derivation of the Bayes estimator for Gaussian prior and Gaussian noise on infinite dimensional spaces. Section 3 contains a discussion of QN-laws defined on locally convex spaces and a description of the posterior distribution when the prior or the noise is a QN-law. Upper bounds for the increase in the mean square error of δ_0 over the minimum possible mean square error under a QN-law prior or QN-law noise are given in Section 4. Some examples, including an application to Kalman filtering, are discussed at the end of the paper.

2. Preliminaries.

Let $(S,\mathcal{S})$ and $(T,\mathcal{T})$ be measurable spaces, μ_{XY} a probability measure on $\mathcal{S} \times \mathcal{T}$, μ_X and μ_Y the projections of μ_{XY}. The conditional distribution $\mu_{X|y}$, if it exists, is defined to be a probability measure on $\mathcal{S}$ for a.e. $d\mu_Y(y)$ such that $\mu_{X|y}(A)$ is measurable as a function of y for each fixed $A \in \mathcal{S}$ and

$$\mu_{XY}(A \times B) = \int_B \mu_{X|y}(A)d\mu_Y(y) \quad \text{for all } A \in \mathcal{S} \text{ and } B \in \mathcal{T}.$$

It follows from the definition that $\mu_{X|y} << \mu_X$ a.e. $d\mu_Y(y)$. The following lemma, which is proved using Fubini's theorem, states the abstract Bayes formula of Kallianpur and Striebel (1968).

Lemma 2.1. Suppose that the conditional distribution $\mu_{Y|x}$ exists and the map $(x,y) \longmapsto \frac{d\mu_{Y|x}}{d\mu_Y}(y)$ is $S \times T$ measurable. Then the conditional distribution $\mu_{X|y}$ exists and

$$\frac{d\mu_{X|y}}{d\mu_X}(x) = \frac{d\mu_{Y|x}}{d\mu_Y}(y) \quad \text{a.e.} \quad d\mu_X \otimes \mu_Y(x,y).$$

The probability measure μ_{XY} is to be defined through a prior distribution μ_X on S and a noise distribution μ_N on T. S is the parameter space and T is the observation space. Let $f: S \times T \to T$ be an $S \times T/T$ measurable map. Define μ_{XY} by $\mu_{XY}(A) = \mu_X \otimes \mu_N \{(x,y) : (x,f(x,y)) \in A\}$. It is easily seen that $\mu_{Y|x}$ exists and is equal to $\mu_N \circ f_x^{-1}$ where $f_x : T \to T$ is defined by $f_x(y) = f(x,y)$. When $\mu_{X|y}$ exists it is called the posterior distribution.

Before going further we need to make a brief detour through the theory of probability measures on topological vector spaces. Let E denote a locally convex topological vector space with topological dual E'. The cylindrical σ-algebra on E is the σ-algebra generated by E' and is denoted $\sigma(E')$. Let μ be a probability measure on $\sigma(E')$ such that $\int_E \langle f,x\rangle^2 \, d\mu(x) < \infty$, for all f in E'. Then μ has a mean m and a covariance operator R and under mild conditions m belongs to E and R maps E' into E (See Vakhania and Tarieladze, 1978). Schwartz (1964) showed that if E is quasi-complete then each covariance operator $R: E' \to E$ has a unique Hilbert space H, which is a vector subspace of E, such that the natural injection j of H into E is continuous and $R = jj^*$. The Hilbert space H is called the reproducing kernel Hilbert space (RKHS) of R. If the RKHS is separable with a CONS

$\{e_n, n \geq 1\}$ then the covariance operator admits a series representation $R = \sum_n je_n \otimes je_n$, where $(u \otimes u)(f) = \langle f,u \rangle u$, for $u \in E, f \in E'$, and the series converges to R in the strong operator topology: $\sum_1^n \langle f, je_n \rangle je_n \to Rf$ in E for all f in E'. A probability measure μ on $\sigma(E')$ is Gaussian if each f in E' is a Gaussian random variable under μ. The methods used in this paper depend on the existence of a separable RKHS for the covariance operators of Gaussian measures. For this reason, we assume throughout that E is quasi-complete and each Gaussian measure μ has a mean $m \in E$, a covariance operator $R:E' \to E$ and a separable RKHS. Such a Gaussian measure is specified by $\mu = N(m,R)$.

Now assume that $\mu_N = N(0,R_N)$ on $\sigma(E')$ with RKHS denoted H_N and injection $j_N:H_N \to E$, $\mu_X = N(m_X,R_X)$ on $\sigma(H_N)$, $(S,\mathcal{S}) = (H_N,\sigma(H_N))$, $(T,\mathcal{T}) = (E,\sigma(E'))$ and $f(x,y) = j_N(x) + y$. Let L_N denote the closure of E' in $L^2(E,\mu_N)$, $U_N:L_N \to H_N$ the unitary operator defined by $U_N f = j_n^* f$, for f in E'. R_X is a trace-class operator on H_N so it has a series representation $R_X = \sum_n \tau_n e_n \otimes e_n$, where $\{e_n, n \geq 1\}$ is a CONS in H_N, $\tau_n \geq 0$ and $\mathrm{tr}(R_X) = \sum \tau_n < \infty$. I denotes the identity operator on H_N.

The following proposition, well known for finite dimensional spaces, gives the posterior distribution $\mu_{X|y}$ for Gaussian prior μ_X and Gaussian noise μ_N.

<u>Proposition 2.2</u>. Let $\mu_N = N(0,R_N)$, $\mu_X = N(m_X,R_X)$. Then the posterior distribution $\mu_{X|y}$ exists as a probability measure on $\sigma(H_N)$ and is given by $\mu_{X|y} = N(m_{X|y}, R_{X|y})$, where

$$m_{X|y} = \sum_n \frac{\tau_n}{1+\tau_n} \{ [U_N^{-1}(e_n)](y) + \frac{\langle e_n, m_X \rangle}{\tau_n} \} e_n, \quad R_{X|y} = R_X(I + R_X)^{-1}.$$

__Proof__. Denote $[U_N^{-1}(e_n)](y)$ by $\alpha_n(y)$. The α_n are i.i.d. $N(0,1)$ random variables under μ_N so that $m_{X|y} \in H_N$ a.e. $d\mu_N(y)$. But, $\mu_N \circ f_x^{-1} \sim \mu_N$ for each $x \in H_N$ (cf. McKeague, 1982, Theorem 2.1) so that by Baker (1976) $\mu_Y \sim \mu_N$. Thus $m_{X|y} \in H_N$ a.e. $d\mu_Y(y)$ and the pair $(m_{X|y}, R_{X|y})$ defines a Gaussian measure on $\sigma(H_N)$ a.e. $d\mu_Y(y)$. Now check the conditions of Lemma 2.1. $\mu_{Y|x}$ exists and is equal to $\mu_N \circ f_x^{-1}$. The map $(x,y) \longmapsto d\mu_{Y|x}/d\mu_Y(y)$ is $\sigma(H_N) \times \sigma(E')$ measurable since

$$\frac{d\mu_{Y|x}}{d\mu_Y}(y) = \frac{d\mu_N \circ f_x^{-1}}{d\mu_N}(y) \quad \frac{d\mu_N}{d\mu_Y}(y)$$

$$= \frac{d\mu_N}{d\mu_Y}(y) \exp \{[U_N^{-1}(x)] - \tfrac{1}{2} \; \| x \|^2_{\mu_N}\}$$

$$= \frac{d\mu_N}{d\mu_Y}(y) \exp \sum \{\alpha_n(y) \; \langle e_n, x\rangle - \tfrac{1}{2} \; \langle e_n, x\rangle^2\},$$

where the Radon-Nikodym derivative $d\mu_N \circ f_x^{-1}/d\mu_N$ is given in McKeague (1982, Theorem 2.1), for instance. Now applying Lemma 2.1, the characteristic functional $\hat{\mu}_{X|y}(u) = \int_{H_N} e^{i\langle u,x\rangle} d\mu_{X|y}(x)$, for $u \in H_N$, as a function of u, is proportional to $\int_{H_N} \lim_{k\to\infty} Z_k(x)\, d\mu_X(x)$, where

$$Z_k(x) = \exp \sum_{n=1}^{k} \{i\langle e_n,u\rangle\langle e_n,x\rangle + \alpha_n(y)\langle e_n,x\rangle - \tfrac{1}{2} \; \langle e_n,x\rangle^2\}.$$

Provided that $\{Z_k, k \geq 1\}$ is uniformly integrable, the result now follows from routine calculations since the $\langle e_n,x\rangle$, $n \geq 1$ are independent $N(\langle e_n,m_X\rangle, \tau_n)$ random variables under μ_X. But

$$\int_{H_N} |Z_k(x)|^2 d\mu_X(x) \leq \int_{H_N} \exp\{2 \sum_{n=1}^{k} \alpha_n(y)\langle e_n, x\rangle\} d\mu_X(x)$$

$$= \exp\{2 \sum_{n=1}^{k} (\alpha_n^2(y)\tau_n + \alpha_n(y)\langle e_n, m_X\rangle)\},$$

which shows that $\{Z_k, k \geq 1\}$ is a.e. $d\mu_Y(y)$ uniformly integrable with respect to μ_X, as required. □

3. QN-Laws.

Let E_1 and E_2 be locally convex spaces. Suppose that $\mu = N(m,R)$ on $\sigma(E_1')$ with RKHS denoted H and injection $j : H \to E_1$; also let $A : E_1 \to E_1'$ be a symmetric non-negative operator, $\alpha \in R$, $a \in E_2$ and $J : E_1 \to E_2$ be a continuous linear map. Provided

$$c^{-1} \equiv \int_{E_1} (\alpha^2 + \langle A(J(x)-a), J(x)-a\rangle) d\mu(x) < \infty,$$

define a probability measure ν on $\sigma(E_1')$ by $\nu = \mu$ if $c^{-1} = 0$, otherwise by the relation

$$\frac{d\nu}{d\mu}(x) = c(\alpha^2 + \langle A(J(x)-a), J(x)-a\rangle).$$

The measure ν is called a QN-law and was introduced on Hilbert space by Gualtierotti (1979). If J^*AJ has a separable RKHS then $c^{-1} < \infty$ if and only if j^*J^*AJj is trace-class, and in this case

$$c^{-1} = \alpha^2 + \mathrm{tr}(j^*J^*AJj) + \langle A(J(m)-a), J(m)-a\rangle.$$

It is always possible to assume that α is either zero or one. We shall assume that $\alpha = 1$ and write $\nu = QN((J,a,A),\mu)$. When $E_1 = E_2$ and J is the identity map write $\nu = QN((a,A),\mu)$. Gualtierotti (1980) calculated the mean and covariance operator of ν for the case of a separable Hilbert space. It is possible to extend this result to separable Banach spaces as follows.

Lemma 3.1. Suppose that E_1 is a separable Banach space and J^*AJ has a separable RKHS. Then the mean m^Q and covariance operator R^Q of ν are given by

$$m^Q = m + u$$
$$R^Q = R + 2cRJ^*AJR - u\otimes u,$$
$$\text{where } u = 2cRJ^*A(J(m)-a).$$

Proof. (Sketch) Assume that $m = 0$ and consider just the evaluation of R_ν. Let $J^*AJ = \sum_n g_n \otimes g_n$, $g_n \in E_1'$. Then, for $f \in E_1'$,

$$\int_{E_1} \langle f,x\rangle^2 \langle AJ(x),J(x)\rangle \, d\mu(x) = \sum_n \int_{E_1} \langle f,x\rangle^2 \langle g_n,x\rangle^2 \, d\mu(x),$$

so that we can reduce to evaluating integrals of the form $\int_{E_1} \langle f,x\rangle^2 \langle g,x\rangle^2 \, d\mu(x)$. Choose $h_n \in E_1'$ such that $j^*(h_n)$, $n \geq 1$ is a CONS for H. Define

$$\pi_k x = \sum_{n=1}^{k} \langle h_n,x\rangle Rh_n, \quad x \in E_1.$$

Then, by Tien (1978, Lemma 2), $\pi_k x$ converges a.s. $[\mu]$ to x. But

$$\int_{E_1} \langle f,\pi_k x\rangle^4 \langle g,\pi_k x\rangle^4 \, d\mu(x) \leq \{\int_{E_1} \langle f,\pi_k x\rangle^8 \, d\mu(x)\}^{\frac{1}{2}} \{\int_{E_1} \langle g,\pi_k x\rangle^8 \, d\mu(x)\}^{\frac{1}{2}}$$
$$\leq 105 \langle Rf,f\rangle^2 \langle Rg,g\rangle^2,$$

since $\langle f,\pi_k x\rangle$ is a $N(0, \sum_{n=1}^{k} \langle Rh_n,f\rangle^2)$ random variable and

$\sum_{n=1}^{k} \langle Rh_n,f\rangle^2 \leq \langle Rf,f\rangle$. It follows that $\{\langle f,\pi_k x\rangle^2 \langle g,\pi_k x\rangle^2, k \geq 1\}$ is uniformly integrable and the Lebesgue convergence theorem can be applied. The integral $\int_{E_1} \langle f,\pi_k x\rangle^2 \langle g,\pi_k x\rangle^2 \, d\mu(x)$ can be calculated using the fact that $\langle h_n,x\rangle$, $n \geq 1$ is an i.i.d. $N(0,1)$ sequence of random variables with respect to m. □

The next proposition shows that the posterior is a QN-law if either the prior is Gaussian and the noise is a QN-law or the prior is a QN-law and the noise is Gaussian. Let $\mu_N = N(0,R_N)$, $\mu_X = N(m_X,R_X)$ as in Section 2 and let $\mu_{X|y}$ denote the corresponding posterior distribution given in Proposition 2.2.

<u>Proposition 3.2</u>. (i) If the prior is $\mu_X = N(m_X,R_X)$ and the noise is $\nu_N = QN((a,A),\mu_N)$ then the posterior is $\nu_{X|y} = QN((j_N,y-a,A),\mu_{X|y})$.

(ii) If the prior is $\nu_X = QN((a,A),\mu_X)$ and the noise is $\mu_N = N(0,R_N)$ then the posterior is $\nu_{X|y} = QN((a,A),\mu_{X|y})$.

The proof of this proposition uses the following consequence of Lemma 2.1.

<u>Lemma 3.3</u>. Let μ_{XY} and ν_{XY} be probability measures on $S \times T$ such that

(a) $\mu_X \sim \nu_X$ and $\mu_Y \sim \nu_Y$;

(b) $\mu_{Y|x}$ and $\nu_{Y|x}$ exist and $\mu_{Y|x} \sim \nu_{Y|x}$ a.e. $d\mu_X(x)$;

(c) the maps $(x,y) \longmapsto d\nu_{Y|x}/d\mu_{Y|x}(y)$, $(x,y) \longmapsto d\mu_{Y|x}/d\mu_Y(y)$ are $S \times T$ measurable. Then $\nu_{X|y}$ exists, $\nu_{X|y} \sim \mu_{X|y}$ a.e. $d\mu_Y(y)$ and

$$\frac{d\nu_{X|y}}{d\mu_{X|y}}(x) = \frac{d\mu_Y}{d\nu_Y}(y)\,\frac{d\nu_{Y|x}}{d\mu_{Y|x}}(y)\,\frac{d\nu_X}{d\mu_X}(x) \quad \text{a.e. } d\mu_X \otimes \mu_Y(x,y).$$

<u>Proof</u>. Using (a) and (b) get

$$\frac{d\nu_{Y|x}}{d\nu_Y}(y) = \frac{d\nu_{Y|x}}{d\mu_{Y|x}}(y)\,\frac{d\mu_{Y|x}}{d\mu_Y}(y)\,\frac{d\mu_Y}{d\nu_Y}(y) \quad \text{a.e. } d\mu_X \otimes \mu_Y(x,y)$$

so that, by (c), the function $(x,y) \longmapsto d\nu_{Y|x} / d\nu_Y(y)$ is $S \times T$ measurable and $\nu_{X|y}$ exists by Lemma 2.1. The proof is completed by applying Bayes formula. □

Proof of Proposition 3.2. (i) $\mu_Y \sim \nu_Y$ since $\mu_{Y|x} \sim \nu_{Y|x}$ for all $x \in H_N$.

$$\frac{d\nu_{Y|x}}{d\mu_{Y|x}}(y) = c_N(1 + \langle A(y-a-j_N x), y-a-j_N x\rangle),$$

so that the map $(x,y) \longmapsto d\nu_{Y|x} / d\mu_{Y|x}(y)$ is $\sigma(H_N) \times \sigma(E')$ measurable. The map $(x,y) \longmapsto d\mu_{Y|x} / d\mu_Y(y)$ is $\sigma(H_N) \times \sigma(E')$ measurable from the proof of Proposition 2.2. Thus, by Lemma 3.3 $\nu_{X|y}$ exists and

$$\frac{d\nu_{X|y}}{d\mu_{X|y}}(x) = \frac{d\mu_Y}{d\nu_Y}(y)\, c_N(1 + \langle A(j_N x-(y-a)), j_N x-(y-a)\rangle),$$

which shows that $\nu_{X|y} = QN((j_N, y-a, A), \mu_{X|y})$. The proof of (ii) is similar. □

4. Bayesian Robustness.

Let δ denote a decision rule for estimating the true signal $x \in H_N$. δ is a measurable function from the observation space E into the parameter space H_N. For prior ν_X and noise ν_N the mean square error of δ is given by

$$r(\nu_X, \nu_N, \delta) = \int_{H_N \times E} \| x-\delta(y) \|^2_{H_N}\, d\nu_{XY}(x,y).$$

The following functions of ν_X and ν_N will be used to measure the robustness of a decision rule δ_0: the increase in the mean square error in using δ_0 over the minimum possible mean square error,

$$\Delta(\nu_X, \nu_N, \delta_0) = r(\nu_X, \nu_N, \delta_0) - \inf_{\delta} r(\nu_X, \nu_N, \delta),$$

and the ratio of the mean square error using δ_0 to the minimum possible mean square error,

$$\Phi(\nu_X,\nu_N,\delta_o) = \frac{r(\nu_X,\nu_N,\delta_o)}{\inf_\delta r(\nu_X,\nu_N,\delta)}.$$

Let δ_o be the optimal (in the mean square sense) estimator for Gaussian prior $\mu_X = N(m_X,R_X)$ and Gaussian noise $\mu_N = N(0,R_N)$. Then $\delta_o(y) = m_{X|y}$, the posterior mean given in Proposition 2.2. The results of this section give some upper bounds on $\Delta(\nu_X,\nu_N,\delta_o)$ and $\Phi(\nu_X,\nu_N,\delta_o)$ for ν_X and ν_N as QN-law contaminations of μ_X and μ_N respectively. First we evaluate the mean square error of δ_o under contaminated prior or contaminated noise. Denote $R_1 = R_{X|y} = R_X(I + R_X)^{-1}$.

<u>Lemma 4.1</u>. (i) Let $\nu_X = QN((a,A),\mu_X)$. Then

$$r(\nu_X,\mu_N,\delta_o) = tr(R_1) + 2c_X tr(AR_1^2),$$

where $c_X^{-1} = 1 + trAR_X + \langle A(m_X-a),m_X-a\rangle$.

(ii) Let $\nu_N = QN((a,A,\mu_N)$. Suppose that E is a separable Banach space and A has a separable RKHS. Then

$$r(\mu_X,\nu_N,\delta_o) = tr(R_1) + 2c_N\, tr(A_N R_1^2),$$

where $A_N = j_N^* A j_N$ and $c_N^{-1} = 1 + tr(A_N) + \langle Aa,a\rangle$.

<u>Proof</u>. (i) $r(\nu_X,\nu_N,\delta_o) = \int_{H_N}\int_E \| m_{X|y} - x\|^2\, d\mu_{Y|x}(y)\,d\nu_X(x)$. But

$$m_{X|y} - x = \sum_{n\geq 1} \frac{\tau_n}{1+\tau_n}\{[U_N^{-1}(e_n)](y) - \langle x,e_n\rangle - \frac{\langle x-m_X,e_n\rangle}{\tau_n}\}e_n,$$

so that

$$\int_E \| m_{X|y}-x\|^2 d\mu_{Y|x}(y) = \sum_{n\geq 1}(\frac{\tau_n}{1+\tau_n})^2\int_E\{[U_N^{-1}(e_n)](y)-\langle x,e_n\rangle - \frac{\langle x-m_X,e_n\rangle}{\tau_n}\}^2 d\mu_{Y|x}(y)$$

$$= \sum_{n\geq 1}(\frac{\tau_n}{1+\tau_n})^2\ (1 + \frac{\langle x-m_X,e_n\rangle^2}{\tau_n^2}),$$

since $[U_N^{-1}(e_n)](y) - \langle e_n, x\rangle$ is a $N(0,1)$ random variable under $\mu_{Y|X}$.

By Lemma 3.1

$\int_{H_N} \langle e_n, x-m_X\rangle^2 d\nu_X(x) = \tau_n + 2c_X\tau_n{}^2\langle Ae_n, e_n\rangle$,

so that

$$r(\nu_X,\mu_N,\delta_\circ) = \sum_{n\geq 1} \left(\frac{\tau_n}{1+\tau_n}\right)^2 \left(1 + \frac{1}{\tau_n} + 2c_X\langle Ae_n, e_n\rangle\right)$$

$$= tr(R_X(I + R_X)^{-1}) + 2c_X tr(AR_X{}^2(I + R_X)^{-2}).$$

(ii) is proved in a similar way. □

The following theorem gives an upper bound on the increase in the mean square error of $\delta_\circ$ over the minimum possible mean square error under a contaminated prior distribution.

<u>Theorem 4.2</u>. Let $\nu_X = QN((a,A),\mu_X)$. Then

$\Delta(\nu_X,\mu_N,\delta_\circ) \leq 4c_1^2\|R_1A\|^2[trR_XR_1 + 2c_X trAR_1^2 + (1+4c_X\|AR_XR_1\|)\|m_X - a\|^2]$,

where $c_1^{-1} = 1 + tr(AR_1)$.

<u>Proof</u>. It is easily checked that $\Delta(\nu_X,\mu_N,\delta_\circ) = \int_E \| m_{X|y} - m^Q_{X|y}\|^2 d\nu_Y(y)$.

By Proposition 3.2 and Lemma 3.1, $m^Q_{X|y} = m_{X|y} + 2c_{X|y} R_{X|y}A(m_{X|y} - a)$, so that

$$\Delta(\nu_X,\mu_N,\delta_\circ) \leq 4c_1^2 \|R_1A\|^2 \int_E \|m_{X|y} - a\|^2 d\nu_Y(y).$$

Now consider

$$\int_E \| m_{X|y} - a \|^2 \, d\nu_Y(y) = \int_{H_N} \int_E \| m_{X|y} - a \|^2 \, d\mu_{Y|x}(y) \, d\nu_X(x).$$

$$\int_E \| m_{X|y} - a \|^2 \, d\mu_{Y|x}(y) = \sum_{n\geq 1} \left(\frac{\tau_n}{1+\tau_n}\right)^2 \int \{[U_N^{-1}(e_n)](y) - \langle e_n, x\rangle + \langle e_n, x-a\rangle + \frac{\langle e_n, m_X - a\rangle}{\tau_n}\}^2 \, d\mu_{Y|x}$$

$$= \sum_{n\geq 1} \left(\frac{\tau_n}{1+\tau_n}\right)^2 \left(1 + \{\langle e_n, x-a\rangle + \frac{\langle e_n, m_X - a\rangle}{\tau_n}\}^2\right).$$

Use Lemma 3.1 to get

$$\int_{H_N} \{\langle e_n, x-a\rangle + \frac{\langle e_n, m_X - a\rangle}{\tau_n}\}^2 d\nu_X(x) =$$

$$\tau_n + 2c_X \langle R_X A R_X e_n, e_n\rangle + 4c_X\left(\frac{1+\tau_n}{\tau_n}\right)\langle e_n, R_X A(m_X - a)\rangle\langle e_n, m_X - a\rangle + \left(\frac{1+\tau_n}{\tau_n}\right)^2 \langle e_n, m_X - a\rangle^2.$$

This yields

$$\int_E \| m_{X|y} - a \|^2 \, d\nu_Y(y) = \sum_{n\geq 1} \{\tau_n^2 (1+\tau_n)^{-1} + 2c_X \left(\frac{\tau_n}{1+\tau_n}\right)^2 \langle R_X A R_X e_n, e_n\rangle$$

$$+ 4c_X \langle A R_X^2 (I+R_X)^{-1} e_n, m_X - a\rangle\langle e_n, m_X - a\rangle + \langle e_n, m_X - a\rangle^2\}$$

$$\leq \operatorname{tr} R_X^2 (I+R_X)^{-1} + 2c_X \operatorname{tr} A R_X^4 (I+R_X)^{-2} + 4c_X \| A R_X^2 (I+R_X)^{-1} \| \, \| m_X - a \|^2$$

$$+ \| m_X - a \|^2,$$

and the result follows. □

It is now possible to give an upper bound on $\Phi(\nu_X, \mu_N, \delta_0)$, and since we are mainly interested in the effects of small amounts of contamination, we state it in the following form.

Corollary 4.3. Let $\nu_X = QN((a,\varepsilon A),\mu_X)$, where $\varepsilon>0$. Then

$$\Phi(\nu_X,\mu_N,\delta_\circ) \le 1 + \frac{4\| R_1A\|^2 \,[\mathrm{tr}(R_XR_1) + \| m_X-a\|^2\,]}{\mathrm{tr}(R_1)}(1 + o(1))\varepsilon^2, \quad \text{as} \quad \varepsilon\to 0.$$

In particular, $\Phi(\nu_X,\mu_N,\delta_\circ) = 1 + O(\varepsilon^2)$, $\varepsilon\to 0$.

Proof. The result follows from Proposition 4.1, Theorem 4.2 and the identity

$$\Phi(\nu_X,\mu_N,\delta_\circ) = 1 + \frac{\Delta(\nu_X,\mu_N,\delta_\circ)}{r(\nu_X,\mu_N,\delta_\circ) - \Delta(\nu_X,\mu_N,\delta_\circ)} \qquad \square$$

The next theorem gives an upper bound on the increase in the mean square error of $\delta_\circ$ over the minimum possible mean square error under a contaminated noise distribution. In order to use the known formulae (Lemma 3.1) for the mean and covariance operator of a QN-law on E it is assumed for the remainder of this section that E is a separable Banach space and A has a separable RKHS.

Theorem 4.4. Let $\nu_N = QN((a,A),\mu_N)$. Then

$$\Delta(\mu_X,\nu_N,\delta_\circ) \le 8c_2^2\{\| R_1A_N\|^2\,[\mathrm{tr}R_XR_1 + 2c_N\mathrm{tr}A_NR_1^2]$$
$$+ \mathrm{tr}\, R_1^2(A_NR_XA_N + A_N^2 + 2c_NA_N^3) + (1 + 4c_N\|A_N\|)\langle AR_NAa,a\rangle\},$$

where $A_N = j_N^*Aj_N$ and $c_2^{-1} = 1 + \mathrm{tr}(A_NR_1)$.

Proof. By Proposition 3.2, $\nu_{X|y} = QN((j_N,y-a,A),\mu_{X|y})$, and by Lemma 3.1, $m^Q_{X|y} = m_{X|y} + 2c_{X|y}\, R_1\, j_N^*A(j_Nm_{X|y}-y + a)$. Thus

$$\Delta(\mu_X,\nu_N,\delta_\circ) = \int_E \| m_{X|y} - m^Q_{X|y} \|^2 \, d\nu_Y(y)$$

$$\leq 4c_2^2 \int_E \| R_1 j_N^* A(j_N m_{X|y} - y + a) \|^2 \, d\nu_Y(y)$$

$$\leq 8c_2^2 \, [\| R_1 A_N \|^2 \int_E \| m_{X|y} - m_X \|^2 \, d\nu_Y(y)$$

$$+ \int_E \| R_1 j_N^* A(j_N m_X - y + a) \|^2 \, d\nu_Y(y)] .$$

It is easily checked that

$$\int_E \| m_{X|y} - m_X \|^2 \, d\nu_Y(y) = tr(R_X R_1) + 2c_N tr(A_N R_1^2).$$

Note that $m_Y^Q = j_N m_X + u$ and $R_Y^Q = j_N R_X j_N^* + R_N + 2c_N R_N A R_N - u \otimes u$,

where $u = -2c_N R_N A(a)$. Hence

$$\int_E \| R_1 j_N^* A(j_N m_X - y + a) \|^2 \, d\nu_Y(y)$$

$$= tr(R_1^2 j_N^* A R_Y^Q A j_N) + \| R_1 j_N^* A(a-u) \|^2$$

$$= tr(R_1^2 (A_N R_X A_N + A_N^2 + 2c_N A_N^3) - \| R_1 j_N^* A(u) \|^2$$

$$+ \| R_1 j_N^* A(a-u) \|^2$$

$$= tr(R_1^2 (A_N R_X A_N + A_N^2 + 2c_N A_N^3) + \| R_1 j_N^* A(a) \|^2$$

$$+ 4c_N \langle R_1 j_N^* A(a), R_1 A_N j_N^* A(a) \rangle$$

$$\leq tr(R_1^2 (A_N R_X A_N + A_N^2 + 2c_N A_N^3) + (1 + 4c_N \| A_N \|) \langle A R_N A a, a \rangle .$$

The result follows immediately. □

Corollary 4.5. Let $\nu_N = QN((a,\varepsilon A),\mu_N)$, where $\varepsilon > 0$. Then

$$\Phi(\mu_X,\nu_N,\delta_\circ) \leq 1 + \frac{8[\| R_1 A_N \|^2 \, tr R_X R_1 + tr R_1^2 (A_N R_X A_N + A_N^2) + \langle A R_N A a, a \rangle]}{tr(R_1)} (1 + o(1)) \varepsilon^2,$$

as $\varepsilon \to 0$. In particular, $\Phi(\mu_X,\nu_N,\delta_\circ) = 1 + O(\varepsilon^2)$, $\varepsilon \to 0$.

Examples

1. *The one-dimensional case with contaminated prior.* Let X and N be independent random variables with distributions $\nu_X = QN((m_X,\varepsilon),\mu_X)$ and $\mu_N = N(0,\sigma_N^2)$ respectively, where $\mu_X = N(m_X,\sigma_X^2)$. Then $Y = X + N$, $A = \varepsilon\sigma_N^2$, and $R_X = \sigma_X^2/\sigma_N^2 \equiv \rho$, the signal to noise ratio. By Corollary 4.3

$$\frac{E\,(X-\delta_\circ(Y))^2}{\inf\limits_{\delta} E(X-\delta(Y))^2} \le 1 + \frac{4\sigma_X^4\rho}{(1+\rho)^2}(1+o(1))\varepsilon^2$$

2. *The one-dimensional case with contaminated noise.* Let X and N be independent random variables with distributions $\mu_X = N(m_X,\sigma_X^2)$ and $\nu_N = QN((0,\varepsilon),\mu_N)$ respectively, where $\mu_N = N(0,\sigma_N^2)$. Then $A_N = \varepsilon\sigma_N^2$, $R_X = \rho$ and by Corollary 4.5

$$\frac{E\,(X-\delta_\circ(Y))^2}{\inf\limits_{\delta} E(X-\delta(Y))^2} \le 1 + 8\sigma_N^4\rho[1 + (\frac{\rho}{1+\rho})^2](1+o(1))\varepsilon^2.$$

3. *Kalman filtering in the presence of contamination.* Let the signal process X_t and the observation process Y_t be given by the stochastic differential equations

$$dX_t = -\beta X_t dt + dW_t^1$$

$$\text{and} \quad dY_t = X_t dt + dW_t^2$$

$(0\le t\le 1)$, where W^1 and W^2 are independent Wiener processes, $\beta>0$, and $X_\circ$ is a $N(0,\frac{1}{2\beta})$ random variable which is independent of W^1 and W^2. Then $E = C[0,1]$, $H_N = L^2[0,1]$, $j_N : H_N \to E$ is defined by $j_N(f)(t) = \int_0^t f(s)ds$, for $f\in H_N$, $t\in[0,1]$, R_N is the integral operator with kernel min (s,t) and R_X is the integral operator on $L^2[0,1]$ with kernel $\frac{1}{2\beta}e^{-\beta|s-t|}$. $\delta_\circ$ can be expressed as the solution of a stochastic differential equation for the interpolation of a Gaussian process (see Liptser and Shiryayev, 1978).

a) Contaminated signal. Let A be the identity operator on $L^2[0,1]$ and let $\nu_X = QN((0,\varepsilon A),\mu_X)$, where $\mu_X = N(0,R_X)$. By Corollary 4.3 $\Phi(\nu_X,\mu_N,\delta_*) \leq 1 + 4\,\mathrm{tr}(R_X)(1+o(1))\varepsilon^2$. But $\mathrm{tr}(R_X) = 1/2\beta$ so that

$$\Phi(\nu_X,\mu_N,\delta_*) \leq 1 + \frac{2}{\beta}(1+o(1))\varepsilon^2.$$

b) Contaminated noise. Let A be the natural injection of $C[0,1]$ into $C^*[0,1]$ and let $\nu_N = QN((0,\varepsilon A),\mu_W)$, where μ_W is Wiener measure on $C[0,1]$. Thus $d\nu_N/d\mu_W(x) = c_N(1 + \varepsilon\int_0^1 x_t^2 dt)$, where c_N is a constant. By Corollary 4.5, $\Phi(\mu_X,\nu_N,\delta_*) \leq 1 + 24\,\mathrm{tr}(R_X)(\mathrm{tr}A_N)^2(1+o(1))\varepsilon^2$. But A_N is the integral operator on $L^2[0,1]$ with kernel min (s,t). Thus $\mathrm{tr}(A_N) = \frac{1}{2}$ and it follows that

$$\Phi(\mu_X,\nu_N,\delta_*) \leq 1 + \frac{3}{\beta}(1+o(1))\varepsilon^2.$$

Acknowledgements. The author would like to thank Debabrata Basu, Roger Berger, Dennis Lindley and Jayaram Sethuraman for some helpful conversations.

REFERENCES

Baker, C.R. (1976). Absolute continuity and applications to information theory. Probability in Banach Spaces, Oberwolfach 1975, Lecture Notes in Mathematics 526, Springer-Verlag, Berlin, 1-11.

Berger, J.O. (1982). The robust Bayesian viewpoint. Technical Report #82-9, Department of Statistics, Purdue University.

Blum, J.R. and Rosenblatt, J. (1967). On partial a priori information in statistical inference. Ann. Math. Statist., 38, 1671-1678.

Box, G.E.P. and Tiao, G.C. (1968). A Bayesian approach to some outlier problems. Biometrika, 55, 119-129.

Ershov, A.A. and Liptser, R.S. (1978). Robust Kalman filter in discrete time. Automation and Remote Control, 39, 359-367.

Gualtierotti, A.F. (1979). Sur la détection des signaux, l'information de Shannon, et la capacité de retransmission, quand les lois du modèle ne sont pas exactement normales. Comptes Rendus Acad. Sci. Paris, A288, 69-71.

Gualtierotti, A.F. (1980). On Average Mutual Information and Capacity for a Channel without Feedback and Contaminated Gaussian Noise. Inform. Contr., 46, 46-70.

Gualtierotti, A.F. (1982). On the stability of signal detection. To appear in IEEE Trans. Information Theory.

Kallianpur, G. and Striebel, C. (1968). Estimation of stochastic processes with additive white noise observation errors. Ann. Math. Statist., 39, 785-801.

Liptser, R.S. and Shiryayev, A.N. (1977). Statistics of Random Processes, Volumes I and II, Springer-Verlag, New York.

Masreliez, C.G. (1975). Approximate non-Gaussian filtering with linear state and observation relations. IEEE Trans. Automat. Contr., 20, 107-110.

McKeague, I.W. (1982). On the capacity of channels with Gaussian and non-Gaussian noise. To appear in Inform. Contr.

Schwartz, L. (1964). Sous-espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés. J. Analyse Math., 13, 115-256.

Solomon, D.L. (1972). A-minimax estimation of a multivariate location parameter. J. Amer. Statist. Assoc., 67, 641-646.

Tien, N.Z. (1979). The structure of measurable linear functionals on Banach spaces with Gaussian measures. Theory Prob. Appl., 24, 165-168.

Vakhania, N.N. and Tarieladze, V.I. (1978). Covariance operators of probability measures in locally convex spaces. Theory Prob. Appl., 23, 1-21.

Watson, S.R. (1974). On Bayesian inference with incompletely specified prior distributions. Biometrika, 64, 193-196.